Article

Hate Speech on Social Media Networks: Towards a Regulatory Framework?

Alkiviadou, Natalie

Alkiviadou, Natalie ORCID: 0000-0002-4159-8710 (2019) Hate Speech on Social Media Networks: Towards a Regulatory Framework? Information and Communications Technology Law, 28 (1). pp. 19-35. ISSN 1360-0834

Hate Speech on Social Media Networks: Towards a Regulatory Framework?

Dr. Natalie Alkiviadou
Lecturer, University of Central Lancashire Cyprus
nalkiviadou@uclan.ac.uk

Abstract

Social networks serve as effective platforms in which users' ideas can be spread in an easy and efficient manner. However, those ideas can be hateful and harmful, some of which may even amount to hate speech. YouTube, Facebook and Twitter have internal regulatory policies in relation to hate speech and have signed a Code of Conduct on the regulation of illegal hate speech with the European Commission. This paper looks at the issue of tackling hate speech on social networks and argues that, notwithstanding the weaknesses of internal policies and their implementation, their existence, as facilitated by the Code of Conduct, serves as a light at the end of the Internet hate tunnel where issues of multiple jurisdictions as well as technological realities, such as mirror sites and more, have resulted in the task of online regulation being more than a daunting one.

Key words: Social media, Hate Speech, Non-discrimination, Internet, Code of Conduct on Illegal Hate Speech

Introduction

Social networks are the frenzy of the 21st century. The latest statistics show that there are 2.19 billion Facebook users,[1] 1.57 billion YouTube users[2] and 336 million Twitter users.[3] Social networks facilitate borderless communication, allow for, *inter alia*, political, ideological, cultural and artistic expression, permit an inflow of daily news, raise awareness on human rights violations and offer a quick and cheap solution to inviting people to your birthday party. At the same time, social networks constitute platforms through which hateful rhetoric is spread[4] and normalised and minority groups are systematically targeted, thereby affecting today's world on a micro (individual), meso (group) and macro (societal) level. Hate existed before the Internet and social networks but the emergence of the Internet and the subsequent creation of social networks have added new dimensions to the already complex topic of hate speech.[5] An important observation needs to be made from the outset, namely the distinction between examining hate on the Internet and examining hate on social networks. The Internet is a global platform which allows for the creation of, amongst others, social networks, news portals and chat rooms. As noted by the Secretary General of the United Nations, Internet use for the objective of promoting hateful expression is one of the most significant human rights challenges that has come about with technological developments.[6] This paper will not look at the regulation of the Internet in its entirety but, instead, focus on the ever powerful tool found on the Internet, namely social networks which, as stated by one commentator, represent 'incredible and unique communication opportunities.'[7] Particular attention needs to be paid to the issue of hate on such networks, rather

[1] Facebook statistics: <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/> [Accessed 12 June 2018]

[2] Youtube statistics: <https://www.omnicoreagency.com/youtube-statistics/> [Accessed 12 June 2018]

[3] Twitter statistics: <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/> [Accessed 12 June 2018]

[4] Fernne Brennan, 'Legislating against Internet Race Hate' (2009) 18 Information & Communications Technology Law 2, 123

[5] Yulia A Timofeeva, 'Hate Speech Online: Restricted or Protected? Comparison of Regulations in the United States and Germany' (2003) 12 Journal of Transnational Law and Policy 2, 255

[6] The Secretary-General, 'Preliminary Representation of the Secretary-General on Globalization and Its Impact on the Full Enjoyment of All Human Rights' paras 26-28, U.N. Doc A/55/342 (Aug 31 2000)

[7] Leandro Silva, Mainack Mondal, Denzil Correa & Fabrício Benevenuto, 'Analyzing the Targets of Hate in Online Social Media' Proceedings of the Tenth International AAAI (Association for the Advancement of Artificial Intelligence) Conference on Web and Social Media (2016) 687

than simply looking at them as part of the general discussion on Internet hate regulation for several reasons. Firstly, the sheer number of users of such networks on a global scale results in the need to pay particular attention to this digital vehicle. Secondly, social networks are used by individual users but also by organised and semi-organised groups to promote hateful rhetoric and target the victims of such rhetoric. Thirdly, social networks come with some kind of content regulation which must be assessed for purposes of ascertaining whether or not and, if so, the extent to which this regulation contributes to the effective tackling of online hate. However, tackling hate on social media is a complex matter with an array of issues that need to be dealt with. Firstly, as is the case with hate speech more generally, there is no universally accepted definition, probably given the fact that 'there is no universal consensus on what is harmful or unsuitable'[8] in this sphere. This means that there cannot be coherence amongst national legal frameworks, which, given the nature of the Internet as a global medium, is necessary if hate speech is to be regulated in an effective manner.[9] Further, determining the best recipe for tackling hate speech on social media is a multi-faceted process. Is it regulation and prohibition? What type of regulation are we contemplating? Is it digital prohibition or criminalisation? Is regulation more generally irrelevant or insufficient? Should we focus on other innovative means to ensure sustainability such as, for example, the promotion of counter-narratives? Is it either or both? The position of the Council of Europe, which is the only institution to draw up a legal document on online (racist and xenophobic) hate is clear. In the Preamble to the Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems, reference is made to the 'risk of misuse or abuse of such computer systems to disseminate racist and xenophobic propaganda' and, although sensitive to the issue of free expression,[10] it was decided, as demonstrated in the title, to criminalise digital acts of racism and xenophobia. It has been argued that the 'harshest' of approaches to online hate speech, namely criminalisation, has been taken by the Council of Europe for fear of such phenomena causing social unrest and damage to the institution's mandate

[8] Irene Nemes, 'Regulating Hate Speech in Cyberspace: Issues of Desirability and Efficacy' (2010) 11 Information and Communications Technology Law 3,195

[9] For analysis of the issue of jurisdiction and online hate regulation look at: Natalie Alkiviadou, 'Regulating Internet Hate: A Flying Pig?' (2016) 7 Journal of Intellectual Property, Information Technology and E-Commerce Law 3

[10] Preamble to the Additional Protocol to the Cybercrime Convention states that the Contracting Parties are 'mindful to the need to ensure a proper balance between freedom of expression and an effective fight against acts of a racist and xenophobic nature.'

which is peace and unity.[11] As such, this paper will look at the issue of hate speech on social networks and the tools available for tackling this speech which, to date, embrace the regulation of such speech through its removal by the networks themselves and through criminalisation. In light of the Code of Conduct agreed upon in 2016 between the European Commission and four IT companies for the regulation of hate speech online, and given that three of those companies constitute the leading social networks of our time, this article will pay particular attention to those, namely Facebook, Twitter and YouTube.

1. The Internet and Social Networks: A New World of Opportunities and Menaces

The Internet is a global giant, 'a network of networks'[12] that, since its creation, has turned physical space and actual time into nothing more than an illusion. It provides 'globalism, anonymity and speed for any on-line activity'[13] and is 'a true marketplace of ideas.'[14] Social networks are digital communities which allow individuals to share messages, images and videos. These have further revolutionised social interaction, expression and the exchange and dissemination of ideas. Such networks function within the digital space of the Internet and allow for quick dissemination of information, synchronous and asynchronous chat, the creation of groups of common interest and ideology. As such, ideas, persons and collectives are placed in a public forum to express, share and mobilise. The effect of online activity on individuals, groups and society, more generally, must not be underestimated. As noted by one commentator, 'today's public consciousness is shaped not in the streets or the parks but in online editorials and web forums.'[15] There, one can find neutral, positive but also harmful information and ideas. Along with the clear advantages that come with social networks, such as access to information and global communication, come perils such as hate speech with such networks constituting ideal

[11] 'A short history of the Council of Europe': <http://www.coe.int/T/E/Com/About_Coe/10_points_intro.asp> [Accessed 1 May 2017]

[12] Nina Vajić & Panayiotis Voyatzis, 'The Internet and Freedom of Expression: A Brave New World and The ECtHR's Evolving Case-Law' in Josep Casadevall, Egbert Myjer, Michael O'Boyle & Anna Austin (eds.), *Freedom of Expression: Essays in Honour of Nicolas Bratza* (Wolf Legal Publishers 2012) 393

[13] Yulia A Timofeeva, 'Hate Speech Online: Restricted or Protected? Comparison of Regulations in the United States and Germany' (2003) 12 Journal of Transnational Law and Policy 2, 254

[14] Candida Harris, Judith Rowbotham & Kim Stevenson, 'Truth, Law and Hate in the Virtual Marketplace of Ideas: Perspectives on the Regulation of Internet Content' (2009) 18 Information & Communications Technology Law 2, 155

[15] Lashel Shaw, 'Hate Speech in Cyberspace: Bitterness without Boundaries' (2012) 25 Notre Dame Journal of Law, Ethics & Public Policy, 280

environments through which this phenomenon can grow. Firstly, the sheer number of users means a large audience while the possibility of pseudonymity,[16] the limitations associated with Internet regulation and, to a lesser but still significant extent, the regulation of material on social networks empower haters to utter their harmful and/or illegal words and share hateful texts, videos and images.[17] There exist no global statistics regarding online hate speech. Nevertheless, key institutions have warned that it is on the rise,[18] while several academic commentators have warned that hate speech is ever present on the Internet and has an ability to cause harm to its targets [19] The Council of Europe Committee of Ministers Declaration on freedom of communication on the Internet,[20] underlined the necessity to ensure freedom of speech and freedom of information, but it also stressed that 'freedom of communication on the Internet should not prejudice the human dignity, human rights and fundamental freedoms of others, especially minors.' However, drawing lines between 'competing' rights and freedoms and working with relatively abstract notions such as harm and dignity is not straightforward. The way such notions have been looked at by law and by regulatory policies of social networks will be discussed in section three below.

When looking at the regulation of online hate on a practical scale, there is a central difference between looking at the Internet in its entirety and looking at social networks in particular. With the former, issues of jurisdiction may arise in relation to, for example, the publication of material which is accessible but illegal in one country but legal in the country hosting the website. Resolving the issue of whether material is impugned or not requires cooperation and agreement between the two countries involved, as was seen in, amongst others, the *Yahoo! Inc. v La Ligue*

[16] UNESCO 'Countering Online Hate Speech' (2015 UNESCO Publishing) 15: 'Genuinely anonymous online communications are rare as they require the user to employ highly technical measures to ensure that he or she cannot be easily identifiable.'
<http://unesdoc.unesco.org/images/0023/002332/233231e.pdf> [Accessed 1 May 2017]
[17] *Ibid.* 14
[18] The proliferation of hate speech online, as observed by the UN Human Rights Council Special Rapporteur on Minority Issues (HRC, 2015) poses a new set of challenges.
[19] As quoted and discussed in Fernne Brennan, 'Legislating against Internet Race Hate' 18 Information & Communications Technology Law 2 (2009) 127
[20] Adopted by the Committee of Ministers on 28th May 2003 at the 840th meeting of the Ministers' Deputies

Contre Le Racism et L'Antisemitisme et al.[21] Regulation of social networks by social networks make the issue of hate on such networks different than hate on many other Internet 'tools.' The three networks looked at in this paper have rules of their own *vis*-à-*vis* prohibited content and, as mentioned in the introduction, a Code of Conduct has been agreed between the European Commission and IT Companies in relation to the IT Companies' role in regulating hate speech. One commentator held that the Internet is 'even worse than a vandalised library because thousands of additional unorganised fragments are added daily by myriad cranks, sages and persons with time on their hands who launch their unfiltered messages into cyberspace.'[22] On one level, that of the Internet in its entirety, this is partly accurate although, given the regulatory policies of social networks and the enhancement of such policies by the Code of Conduct, material may be unfiltered to begin with but the possibility of it being removed does exist.

2. Hate Speech: Meaning and Effects

Before proceeding to consider the mechanisms which do or may exist in terms of tackling online hate, it is important to look at the meaning of hate speech and its potential consequences.

2(i) Hate speech meaning

There is no universally accepted definition of hate speech.[23] This is a result of two main reasons, the varying interpretation of free speech, predominantly between countries or regions, and the interlinked differentiations in the conceptualisation of harm. As argued, hate speech lies 'in a complex nexus with freedom of expression and group rights, as well as concepts of dignity, liberty and equality.'[24] On a European level, hate speech needs to go beyond types of expression that 'shock, offend or disturb'[25] which, according to the European Court of Human Rights (ECtHR), fall within the protective scope of Article 10 of the European Convention on Human

[21] Yahoo, Inc. v. La Ligue Contre Le Racisme et L'Antisemitisme, et al 145 F. Supp. 2d 1168, Case No. C-00-21275JF (N.D. Ca., September 24, 2001). For more info look at Natalie Alkiviadou, 'Regulating Internet Hate: A Flying Pig?' (2016) 7 Journal of Intellectual Property, Information Technology and E-Commerce Law 3

[22] Patrick J. Fahy, '*Achieving Quality with Online Teaching Technologies*' (eds. 2000 ERIC Clearinghouse) 13

[23] General Recommendation No. 32 on The Meaning and Scope of Special Measures in the International Convention on the Elimination of Racial Discrimination (2009) CERD/C/GC/32, para.9

[24] Leandro Silva, Mainack Mondal, Denzil Correa & Fabrício Benevenuto, 'Analyzing the Targets of Hate in Online Social Media' Proceedings of the Tenth International AAAI (Association for the Advancement of Artificial Intelligence) Conference on Web and Social Media (2016) 688

[25] As stated by the ECtHR in Handyside v UK, Application no. 5493/72 (ECHR 1976)

Rights. One of the few, albeit non-binding documents, which defines hate speech is the Recommendation of the Council of Europe Committee of Ministers on hate speech.[26] According to the Ministers, this term is to be:

> 'understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin.'

Although broad in the sense that it includes the justification of hatred as a form of hate speech, it is narrow in terms of content. More specifically, where is the protection of groups such as LGBTI (Lesbian, Gay, Bisexual, Transgender and Intersex) and disabled persons? This is a question that comes up in several documents, binding and non-binding, which seek to tackle the issue of hate, discussed in section three below. The Additional Protocol to the Convention on Cybercrime deals only with acts of a racist and xenophobic nature, the Framework Decision of the European Union chose to deal only with Racism and Xenophobia while there is no counterpart of the International Convention on the Elimination of All Forms of Racial Discrimination that opted to deal with the protection of victims of homophobic, biphobic or transphobic speech. Whilst the Disability Convention does exist, it does not tackle the issue of hate speech. Since international human rights law contends that all humans are born free and equal in dignity and in rights, [27]why should only racist and xenophobic speech be prohibited by an international document? This malaise in the regulatory framework has led to what can be referred to as a hierarchy of hate, where protection against hate speech is granted to victims of only some 'genres' of hate speech. The policies and procedures of two out of the three social networks discussed in this paper opt for a much broader definition of hate speech which does not omit entire groups which are significantly and systematically victims of hate speech (and not only), namely LGBTI persons, as well as other groups including, but not limited to, disabled persons. Given the severity, in human rights terms, of ignoring marginalised groups such as LGBTI persons and openly prioritising a certain type of hate speech, this differentiation cannot

[26] Council of Europe's Committee of Ministers Recommendation 97 (20)
[27] Article 1, Universal Declaration of Human Rights

be justified. Documents such as the Framework Decision and the Additional Protocol result in criminal penalties whereas the policy, terms and conditions of a social network site will result in the removal of a post or, in the worst case scenario, the banning of a user. So, the community guidelines and terms of two out of the three largest social networks do not ignore groups such as LGBTI persons. However, the effects of their regulatory action are softer than the effects of, for example, the implementation of a national law transposing the EU's Framework Decision on Racism and Xenophobia.

2 (ii) Hate speech definition by social media

Several social media sites provide their own definition of hate speech. YouTube's terms of service directly refer to hate speech, stating that:

> 'we encourage free speech and defend everyone's right to express unpopular points of view. But we do not permit hate speech: speech which attacks or demeans a group based on race or ethnic origin, religion, disability, gender, age, veteran status and sexual orientation/gender identity.'[28]

Facebook community standards refer directly to the removal of hate speech, defining it as:

> 'content that directly attacks people based on their race, ethnicity, national origin, religious affiliation, sexual orientation, sex, gender or gender identity or serious disabilities or diseases.' Further, organisations and people who are dedicated to 'promoting hatred against these protected groups are not allowed a presence on Facebook.'[29]

Twitter does not refer to hate speech but, instead, warns (rather than prohibits) users that they may be exposed to content that might be 'offensive, harmful, inaccurate or otherwise inappropriate…' Importantly, its terms provide that it 'may not monitor or control the Content

[28] YouTube's Community Guidelines: <https://www.youtube.com/yt/policyandsafety/communityguidelines.html> [Accessed 1 May 2017]
[29] Facebook's Community Guidelines: <https://www.facebook.com/communitystandards#hate-speech> [Accessed 2 May 2017]

posted via the Services and, we cannot take responsibility for such Content.'[30] The only prohibition is that of 'direct, specific threats of violence against others.'[31] The case of Twitter is a paradox given that, as will be discussed later on, it is part of the Code of Conduct on Countering Illegal Hate Speech Online which requires the IT Companies to, *inter alia*, remove such speech within 24 hours of receiving a report. Therefore, both YouTube and Facebook include a larger sphere of potential victims of hate speech than key documents such as the International Convention on the Elimination of All Forms of Racial Discrimination, the Framework Decision on Racism and Xenophobia or the Additional Protocol to the Cybercrime Convention do. In addition, Facebook prohibits speech which 'attacks' people based on the aforementioned characteristics and YouTube prohibits speech which 'attacks or demeans' a group. Three issues arise here. Firstly, YouTube has the widest scope of prohibited activity as it also incorporates the demeaning of persons. Secondly, notwithstanding the widest scope, YouTube refers to the prohibition of speech against a certain group rather than a person belonging to that group. Does this mean that if a particular expression were directed against an individual who belonged to a group because of that group's characteristics (rather than against the group in its entirety), such speech would not be prohibited? Given the general content of the Code of Conduct between the European Commission and the IT Companies elaborated on below, it is more likely that it is an issue of the wrong use of language. Thirdly, notwithstanding the Code of Conduct, Twitter seems to limit the control of speech on its network, unless this amounts to a direct and specific (rather than an abstract and generalised) threat of violence. This approach is stricter than the legal documents referred to above which bring about criminal penalties.

3. Hate Speech on Social Networks: Regulatory Framework

3(i) Harm as a necessary pre-requisite for regulation

The issue of hate speech regulation is usually presented by academics, civil society and international organisations as a balancing exercise between free speech and other freedoms and values such as freedom of discrimination and human dignity. In relation to freedom of expression, it must be noted that this has been a key concern for many in relation to hate speech regulation. Countries have incorporated reservations to Article 4 of the International Convention

[30] Twitter's Terms of Service (Content): <https://twitter.com/tos?lang=en#usContent> [Accessed 1 May 2017]
[31] ibid

on the Elimination of All Forms of Discrimination on grounds of free speech [32]and both the Framework Decision on Racism and Xenophobia and the Additional Protocol to the Cybercrime Convention, discussed below, include a 'safety net' of free speech. The Framework Decision holds that it 'shall not have the effect of modifying the obligation to respect fundamental rights and fundamental legal principles, including freedom of expression.'[33] The Additional Protocol is 'mindful of the need to ensure a proper balance between freedom of expression and an effective fight against acts of a racist and xenophobic nature.'[34] One commentator argues, in relation to the Additional Protocol, an argumentation which can be extended to the Framework Decision and the current approach to hate speech regulation, more generally, that free speech is 'the sacred cow against which the legislation seeks to justify its apparent encroachment for the sake of providing a measure to prohibit cybercrimes motivated by race hate.'[35] Either way, this is the *modus operandi* at a national and international level in relation to hate speech regulation. As such, the balancing exercise needs to be determined. In doing this, the harm resulting from hate speech needs to be determined. Tsesis, a supporter of hate speech regulation, argues that 'prejudicial speech initiates, perpetuates and aggravates socially accepted misrepresentation about outgroups.'[36] As well as harming the targets, it has been argued that hate speech can also affect the perpetrator who is 'more likely to become entrenched in his or her hateful beliefs if given the legitimacy of a global audience.'[37] Victims may suffer emotions such as 'sadness, pain, distress'[38] as well as, amongst others, 'humiliation, isolation and dignitary affront.'[39] However, are these consequences sufficient to limit the fundamental right to free speech? A definitive, universally accepted answer cannot be given here or anywhere, for three reasons. Firstly, there

[32] Countries such as Belgium and Austria: <https://treaties.un.org/Pages/ViewDetails.aspx?src=IND&mtdsg_no=IV-2&chapter=4&lang=en#EndDec> [Accessed 2 May 2017]

[33] Article 7, Framework Decision on combating certain forms and expressions of racism and xenophobia by means of criminal law

[34] Preamble, Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems

[35] Fernne Brennan, 'Legislating against Internet Race Hate' (2009) 18 Information and Communications Technology Law 2, 124.

[36] Alexander Tsesis, '*Destructive Messages: How Hate Speech Paves the Way for Harmful Social Movements'* (eds. NUY Press 2002) pg.138

[37] Lashel Shaw, 'Hate Speech in Cyberspace: Bitterness without Boundaries' (2012) 25 Notre Dame Journal of Law, Ethics & Public Policy, 282

[38] Friedrich Kubler, 'How Much Freedom for Racist Speech? Transnational Aspects of a Conflict of Human Rights' (1998) 27 Hofstra Law Review 2, 335

[39] The link between dignity and hate speech has been made by authors such as Richard Abel in '*Speaking Respect, Respecting Speech'* (eds. University of Chicago Press 1998)

are varying approaches to the importance and meaning of free speech as is, for example, classically witnessed in contrasts between US jurisprudence and European jurisprudence on hate speech cases. Secondly, it could be argued that 'hate speech seems to be whatever people choose it to mean.'[40] This is reflected in the different stances taken by international treaties and also the varying definitions put forth by social networks. Thirdly, there are different conceptualisations of what harm actually is and what threshold of harm needs to be established if free speech is to be curtailed. While some commentators, such as Tsesis, emphasise the harm in hate speech and argue in favour of its regulation, others, such as Post, promote non-regulation in the name of the democratic significance of public discussion. As such, he argues that 'racist speech is and ought to be immune from regulation.'[41] Given the theoretical backdrop of his line of argumentation, one could assume that he extends this position to other types of speech such as homophobic and transphobic speech. Others, such as Wright, argue that stringent regulation could be destructive for the victims themselves as this would promote paternalistic attitudes towards institutionally perceived minorities.[42] The issue of paternalism, but through another lens, was also raised by the UN Special Rapporteur on Freedom of Expression who held that excessive regulation of the Internet in order to 'preserve the moral fabric and cultural identity of societies is paternalistic.'[43] What is of particular interest is that the aforementioned balancing exercise and the long-standing debate of free speech versus hate speech, which have led not only to academic discussion on the matter but also to cases such as *Yahoo!* mentioned above and the decision of the USA not to sign the Additional Protocol to the Cybercrime Convention, do not actually affect the regulation (non criminalisation) of hate speech on the three social networks discussed in this paper. This is because Facebook and YouTube have their own terms and conditions which directly prohibit hate speech and Twitter, although it absolves itself of responsibility in relation to content on the network, has signed the Code of Conduct on hate speech and has, thus, committed to removing hate speech content. As such, and as will become apparent in this section, social networks

[40] Roger Kiska, 'Hate Speech: A Comparison Between The European Court of Human Rights and the United States Supreme Court Jurisprudence' (2012) 25 Regent University Law Review 107, 110

[41] Robert C.Post, 'Racist speech, Democracy and the First Amendment' (1991) 32 William and Mary Law Review 267, 322

[42] George Wright, 'Dignity and Conflicts of Constitutional Values: The Case of Free speech and Equal Protection' (2006) 43 San Diego Law Review 527, 566

[43] Report of the Special Rapporteur, Mr. Abid Hussain, submitted pursuant to Commission on Human Rights resolution 1997/26 (28 January 1998) E/CN.4/1998/40, para. 45.

discussed in this paper could, in fact, be the easiest of Internet 'platforms' in terms of hate speech regulation if regulation is what we are ultimately seeking. Also, regulation essentially means removing impugned material and, in cases of systematic misbehaviour by the user, potentially warning or banning the user from the network. This is not a prison sentence as is, for example, incorporated in the EU's Framework Decision. Instead, the regulation by social networks is less harsh and could, therefore, be a middle ground between free speech absolutists and those wanting hate speech regulation even for types of speech which are not direct calls to violence.

3(ii) Legal regulation of online hate

The International Convention on the Elimination of All Forms of Discrimination does not address the issue of online hate directly but, instead, prohibits certain types of expression. More particularly, it holds that States 'shall declare an offence punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin…'[44] In discussing this provision in *Gelle v Denmark*, the Committee on the Elimination of All Forms of Racial Discrimination observed that:

> 'it does not suffice, for purposes of Article 4 of the Convention, merely to declare acts of racial discrimination punishable on paper. Rather, criminal laws and other legal provisions prohibiting racial discrimination must also be effectively implemented by the competent national tribunals and other State institutions. This obligation is implicit in Article 4 of the Convention.'[45]

Article 20(2) of the International Covenant on Civil and Political Rights provides that 'any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.' The above two documents were designed in a pre-Internet era and, thus, online hate was not a consideration. However, the provisions can be

[44] Article 4 (a), International Convention on the Elimination of All Forms of Racial Discrimination

[45] Gelle v Denmark, Communication no. 34/2004 (15 March 2006) CERD/C/68/D/34/2004, para. 7.3. This was reiterated in Jama v Denmark, Adan v Denmark and TBB-Turkish Union v Germany.

used to regulate hate speech found on the Internet as it is not the objective and effects of the phenomenon that has changed but, rather, the vehicle it uses for dissemination. The downside of these provisions is that, probably due to socio-historical reasons at the time of drafting, these provisions only tackle racist and religiously discriminatory speech. The threshold of the two documents is similar, talking of hatred, discrimination and violence as a result of the impugned speech, with the International Convention on the Elimination of All Forms of Racial Discrimination incorporating the prohibition of ideas of racial superiority.

On a European Union level, the central document that can be used for the criminalisation of hate speech is the Framework Decision on combatting certain forms and expressions of racism and xenophobia by means of criminal law.[46] Although this document does not directly define hate speech, it prohibits different forms of expression and acts that fall within the framework of "Offences Concerning Racism and Xenophobia." Further, this document does not tackle the issue of online activity but neither does it exclude it. Article 1, therein, entitled 'offences' concerning racism and xenophobia holds that:

> 1. Each Member State shall take the measures necessary to ensure that the following intentional conduct is punishable:
> (a) publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin; and
> (b) the commission of an act referred to in point (a) by public dissemination or distribution of tracts, pictures or other material.

It also includes two provisions on the prohibition of publicly condoning, denying or grossly trivialising crimes of genocide, crimes against humanity and war crimes when the conduct is carried out in a manner likely to incite to violence or hatred against such a group or a member of such a group. In relation to Article 1, the Framework Decision states that Member States may choose only to punish conduct which is either carried out in a manner likely to disturb public order or which is threatening, abusive or insulting. This provision serves as a tool for States that

[46] Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law

wish to limit the scope of expression that falls within the Framework Decision. Article 3 directly stipulates that the prohibition of these acts needs to happen through criminal law and sets out penalties which should be between 1 and 3 years of imprisonment.

The only document that has been designed solely for the purpose of online hateful activity is the Council of Europe's Additional Protocol to the Convention on Cybercrime, concerning the ciminalisation of acts of a racist and xenophobic nature committed through computer systems which was adopted in 2003 and entered into force in 2006. This tool deals with racism and xenophobia in the framework of online threats, insults, materials and genocide denial. As is the case with the Framework Decision, this document includes tools for Member States that wish to limit the scope of its application. For example, apart from racist and xenophobic threats, parties have the possibility of incorporating a provision that punishment will only occur if an act leads to hatred, contempt or ridicule (racist and xenophobic insult), may choose not to criminalise conduct if other effective remedies are available (racist and xenophobic material) or may even choose not to apply a provision (for example in relation to insults and denial, gross minimisation, approval or justification of genocide or crimes against humanity).

The issue of intent is particularly important in relation to the Additional Protocol as all the offences, therein, require that they occur with intent. In the Explanatory Report to the Protocol, it is held that:

> 'the exact meaning of intentionally should be left to national interpretation…It is not sufficient for example for a service provider to be held criminally liable under this provision, that such a service provider served as a conduit for, or hosted a website or newsroom containing such material, without the required intent under domestic law in the particular case. Moreover, a service provider is not required to monitor conduct to avoid criminal liability.'[47]

[47] Explanatory Report to the Additional Protocol to the Cybercrime Convention, Para 25

So, the Protocol seeks to limit liability of, for example, Internet Service Providers (ISPs) which had no intent for impugned material to be disseminated through their service. However, it leaves the interpretation of intent to be a question of national law. The fallibility of the intervention of ISPs in relation to hate speech was manifested in the request from Germany to Deutsche Telekom to prevent user access to the website of the revisionist Ernst Zündel. Although Deutsche Telekom accepted this request, users in the USA made the website's content available to German users through mirror sites. Therefore, this reflects that even if ISPs restrict available content, there are ways of overcoming this restriction and making material available again. [48]

As noted in section two of this paper, the European Union and the Council of Europe, in their respective documents, chose to focus solely on the criminalisation of racism and xenophobia, disregarding other phenomena which are present in the region today such as homophobia, biphobia and transphobia. Although some justification can be granted to older United Nations documents which opted to look solely at race and religion, given the particular social and historic contexts in which they were drafted, an analogous justification cannot be found with the Framework Decision which was passed in 2008. This reality further reinforces the argument that a hierarchy of hate exists.

On a Council of Europe level, in the landmark case of *Delfi v. Estonia*[49], the ECtHR ruled that Internet intermediaries should remove defamatory comments against individuals and that Internet news portals may be liable for offensive commentary made available thereon. This case did not deal with social media, *per se*, but the central principles developed by the ECtHR therein could be transferable to a social media setting. More particularly, the applicant company was the owner of Delfi, one of the largest Internet news portals of Estonia. In 2006, it published an article entitled 'SLK Destroyed Planned Ice Road.' SLK was the abbreviation of a shipping company and L. was a member of its board and the sole shareholder at the time. The posting of the article led to 185 comments, 20 of which contained personal threats and offensive language against L. The ECtHR held that by finding Delfi liable for the defamatory comments, domestic courts were not in breach of Article 10, given that the comments were insulting and threatening, the portal

[48] James Bank, 'Regulating Hate Speech Online' (2010) 24 Computers & Technology 3, 281
[49] Delfi AS v Estonia, App. No 64569/09 (ECHR 16 June 2015)

was professionally managed and commercial and the measures taken by the portal to avoid damage to L were insufficient.[50] The emphasis placed by the ECtHR on the importance of tackling hate speech is significant for the present discussion. More particularly, it held that:

> '…where third-party user comments are in the form of hate speech and direct threats to the physical integrity of individuals, the member States may be entitled to impose liability on Internet news portals if they fail to take measures to remove clearly unlawful comments without delay, even without notice from the alleged victim or from third parties.'[51]

As well as placing importance on the issue of hate speech, the ECtHR also underlined the degree of control that Delfi had on what was made available on the portal. More particularly, it held that Delfi exercised 'a substantial degree of control over the comments published on its portal'[52] and that because it was part of the process of making public the comments on the portal, it 'went beyond that of a passive, purely technical service provider.'[53] Here, a parallel can be drawn with social media platforms including, amongst others, Facebook and Youtube which are not merely technical but are part of the process of the publishing of third party commentary through a notice and take down system like in Delfi and in the case discussed immediately below. In line with the ECtHR's judgement, social media platforms could, therefore, be considered to have substantive control over the comments published on them, a point that positively correlates, as per Strasbourg's view, with a duty on removing material when this contains hate speech and threats.

In 2016, the ECtHR passed another judgement in relation to the intermediary liability of the Internet. The applicants were two Hungarian websites, MTE and INDEX. MTE was a self-regulatory body of Hungarian Internet content providers and INDEX was a large Hungarian news portal. Both allowed user generated commentary. In 2010, MTE published an opinion about two real estate management websites. Later on, INDEX reproduced the opinion. Comments were published by users against the estate managements, both on MTE's website and on INDEX's portal. However, in this case, the Court found a violation of Article 10 and

[50] ibid para.156
[51] ibid para.100
[52] ibid para.153
[53] ibid para.146

differentiated it from *Delfi* by noting that the comments in the case against Hungary were 'notably devoid of the pivotal element of hate speech'[54] whilst MTE was a regulatory rather than a commercial body, and its professional nature 'was unlikely to provoke heated discussions on the Internet.'[55] Of paramount importance is the emphasis placed by the Court on the existence of hate speech amongst the comments as a central indicator for a non-violation of Article 10. Moreover, in both cases, the applicants had a notice and take down system. In *Delfi,* however, this was deemed insufficient as it allowed the material to remain publicly available for six weeks, causing damage to the individual targeted.[56] In *MTE,* the Court found that such a system was a good way to balance conflicting rights and denoted the lower threshold of efficacy of such a system if hate speech was not part of the commentary. [57] The implication in *Delfi* on the efficacy of a control system is significant in the ambit of an online hate discussion insofar as it imposes a strict responsibility on news portals rigorously to monitor and swiftly take down hate speech. Following this line, a notice and take down procedure on social media in itself does not demonstrate efficiency and sufficiency; it also needs to be quick so as to limit the harm done on the targeted person or persons. This is anyhow set out by the Code of Conduct, which requires IT companies to review report material within 24 hours.

3(iii) Towards enhancing the legal regulation of hate speech on social media in the European Union

A relatively recent step taken on a European Union level to enhance the effectiveness of the legal regulation of hate speech on social media is the proposed amendment to the Audiovisual Media Services Directive.[58] The proposal was adopted by the European Commission in 2016 and incorporates, *inter alia*, provisions to prohibit hate speech for purposes of aligning the Directive with the Framework Decision on Racism and Xenophobia. [59] In this sense, the Directive will

[54] Magyar Tartalomszol Galtatok Egyesulete and Index.hu ZRT v Hungary, App. No 22947/13 (2 May 2016) Para. 70
[55] ibid para.73
[56] ibid para.152
[57] ibid para. 91
[58] Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive)
[59] European Commission: Revision of the Audiovisual Media Services Directive: < https://ec.europa.eu/digital-single-market/en/revision-audiovisual-media-services-directive-avmsd> [Accessed 15 October 2017]

prohibit the transfer of material that incites to violence and hatred directed against a group of persons or a member of such a group defined by reference to sex, race, colour, religion, descent or national or ethnic origin. In addition to the Framework Decision on Racism and Xenophobia tackling hate speech, more generally, and the Code of Conduct agreed between the four IT companies and the European Commission, this step demonstrates the severity which the European Union seems to be slowly attaching to combatting hate speech and online hate speech.

3(iv) Regulation by social networks: The Code of Conduct

The social networks discussed in this paper have adopted certain standards and procedures in relation to prohibited expression and material, which appear thereon. More particularly, if a Facebook user finds material or expression on the network to constitute hate speech, as defined by this network, he or she can report it to Facebook. Facebook will then review it and consider whether or not it does in fact constitute hate speech (according to the opinion of the particular handler). If it does, it is then removed. Given that hate speech prohibition falls within the general Community Standards of Facebook, the consequences of violating these standards vary. As well as the removal of material, in the case of systematic misuse of the network, users may receive a warning, restriction of activity on Facebook or even a ban.[60] The same reporting process exists for both YouTube and Twitter. With a view to cracking down on hate speech on social networks, the Code of Conduct on countering illegal hate speech online (31 May 2016) was signed between the European Commission and four IT Companies, namely Facebook, Microsoft, YouTube and Twitter. The Code of Conduct underlined the importance of the companies in relation to the promotion of free speech as well as their commitment to tackling illegal hate speech, as defined by the Framework Decision on Racism and Xenophobia. Key commitments incorporated into the code of conduct include the following:

> Upon receipt of a valid removal notification, the IT Companies are to review such requests against their rules and community guidelines and, where necessary, national laws transposing the Framework Decision 2008/913/JHA, with dedicated terms reviewing requests.

[60] Facebook's Community Guidelines: https://www.facebook.com/communitystandards#hate-speech [Accessed 2 May 2017]

The IT Companies are to review the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary.

At the end of 2016, the European Commission conducted the first monitoring exercise to ascertain whether the social networks (YouTube, Twitter and Facebook) were doing what had been agreed in the Code of Conduct. For a period of six weeks, a total of twelve civil society organisations from nine different Member States reported a total of six hundred cases of what they considered to be illegal hate speech online to the IT Companies and recorded the responses, actions and timing of responses and actions. Out of the six hundred notifications, in one hundred and sixty nine cases (28.2%) the content was removed. Facebook removed the content in 28.3% of cases, Twitter in 19.1% and YouTube in 48.5%. Further, in 40% of the cases, IT Companies reviewed the notification on the same day and in 43% on the day after. Here, it must be reiterated that the Code of Conduct requests that content is dealt with within 24 hours, something which evidently worked in less than half of the cases. Also, YouTube and Twitter were keener to remove cases reported by trusted reporters.[61] This status is given to particular individuals who are, for example, members of organisations with a particular expertise on the issue of hate speech and are, thus, considered to be trusted as reporters (at least more trusted than regular users). However, the vast majority of users of social networks are normal users who still come across hate speech and wish to report it. The over-reliance on users is a large obstacle in relation to hate speech regulation on social networks. By bringing in 'trust issues' of regular network users, as was demonstrated in the first monitoring cycle described above, YouTube and Twitter reduced the effectiveness of the process. The second monitoring period was launched in March 2017 and lasted for a period of seven weeks. An evaluation was carried out by NGOs and public bodies in a total of twenty-four member states. This monitoring exercise reflected that the social networks made significant progress in terms of their commitments under the Code of Conduct. For example, social networks removed 59% of the reported content which was more than double the percentage of the previous monitoring period. Further, the amount of notifications reviewed

[61] Youtube: Removal of 29% of content if it was a normal user and 68% if it was trusted user: Twitter: Removal of 5% of content if it was a normal user and 33% if it was a trusted user: <http://webcache.googleusercontent.com/search?q=cache:VckMt2f4jiEJ:ec.europa.eu/newsroom/document.cfm%3Fdoc_id%3D40573+&cd=1&hl=en&ct=clnk&gl=cy> [Accessed 2 May 2017]

within 24 hours improved from 40% to 51%.[62] A third monitoring exercise commenced in November 2017 and was completed by December 2017. On average, IT companies removed 70% of the prohibited content and in 81.7% of the cases did so in less than 24 hours. By the third monitoring period, the trusted flagger issue, as described above, improved as the only issue determined was that of feedback. More particularly, Twitter and Youtube provided more feedback to trusted flaggers than to normal users.[63]

The process of reporting hate speech on social networks has been described by one commentator (in the context of YouTube in particular) as an 'over policing of hate speech that unfairly infringes on users' freedom of expression.'[64] It is argued that the network handlers responsible for reviewing reported material may remove it 'merely because they disagree with the viewpoint of the speaker, no matter how appropriate others might find the content.'[65] Although this commentator refers to past incidences where YouTube has been found to restrict expression arbitrarily, the results of the first monitoring period of the Code of Conduct discussed below demonstrate that, in most cases, expression reported as hate speech by civil society organisations was not readily removed by the IT companies. In discussing the alleged over-policing, it was argued that, instead of a private review system, a user objecting to material should publically comment on it for purposes of allowing public discussion which the handler of the network could take into account and make a public decision which he/she explains. This is argued on the basis of transparency and involvement of the local community in removing hate speech from social networks.[66] In theory, this might seem like a good idea but, in practice, there are too many obstacles to allow it to materialise effectively. These could include the potential wish for the reporting user to remain anonymous, the endless and insubstantial commentary that could come under a public report, the use of the commentary by haters to promote their speech further and

[62] European Commission Press Release: Countering online hate speech: Commission initiative with social media platforms and civil society shows progress (1 June 2017) <http://europa.eu/rapid/press-release_IP-17-1471_en.htm>
[63] European Commission Fact Sheet: Code of Conduct on countering illegal hate speech online: Results of the 3rd monitoring exercise (January 2018) <http://webcache.googleusercontent.com/search?q=cache:OQDS0SZGbYAJ:ec.europa.eu/newsroom/just/document.cfm%25253Fdoc_id%25253D49286+&cd=1&hl=en&ct=clnk&gl=cy> [Accessed 12 June 2018]
[64] Lashel Shaw, 'Hate Speech in Cyberspace: Bitterness without Boundaries' (2012) 25 Notre Dame Journal of Law, Ethics & Public Policy, 299
[65] ibid
[66] ibid 302

the sheer amount of material on social networks, particularly those such as YouTube, Facebook and Twitter that operate on a global scale, which makes considering even private reports a tricky issue, let alone a public report and an array of comments. Also, as argued by Oboler, the CEO of the Online Hate Prevention Institute[67], 'the longer the content stays available, the more damage it can inflict on the victims and empower the perpetrators.'[68] Although an argument against this would be the positive impact of counter-narratives expressed on commentary under the impugned material, this is a generalisation as it emanates from the faulty premise that the comments will be structured, positive and serve as effective counter narratives. As such, while facilitating public discussion of what is acceptable and what is not and promoting the significance of counter-narratives to hate and haters, this should be kept separate from the technical process of hate speech reporting. Although I do not consider that bias and personal views are an issue of concern in terms of the current reporting process, particularly given the results of the first monitoring period of the Code of Conduct, one of the central obstacles is that social networks predominantly rely on users to report hate speech. As a result, the problem of enforcing the Code of Conduct is a real one given that its operation is absolutely reliant on the users of social media to come across, identify and report hate speech.

3(v) Enhancing regulation by social networks: The Network Enforcement Act

In June 2017, the German Parliament passed an Act to Improve the Enforcement of Rights on Social Networks (The Network Enforcement Act). This is a German Law which applies to social media networks which have two or more million users,[69] such as Facebook, Youtube and Twitter. The law obliges the networks to, *inter alia,* remove 'clearly illegal' content within 24 hours[70] after receiving a user notification, echoing the time frame of the Code of Conduct. In the event that content is not clearly illegal, social networks have seven days to review and remove the content.[71] In determining the illegality of content, the Act refers to the provisions of the German Criminal Code on, amongst others, the dissemination of propaganda material or use of symbols

[67] Online Hate Prevention Institute: http://ohpi.org.au/ [Accessed 12 June 2018]
[68] UNESCO 'Countering Online Hate Speech' (2015 UNESCO Publishing): <http://unesdoc.unesco.org/images/0023/002332/233231e.pdf> [Accessed 1 May 2017]
[69] Article 1.1(2) of the Act
[70] Article 3.2 (2) of the Act
[71] The seven-day deadline may be extended if the social network hires an outside agency to perform the vetting process, a recognized Agency of Regulated Self-Regulation.

of unconstitutional organisations, the encouragement of the commission of a serious violent offence endangering the state, the public incitement to crime and the incitement to hatred. Every social media network that falls within the scope of this piece of legislation is required to appoint a domestic agent to facilitate the process and respond to competent authorities. The Act includes further obligations on the social media networks, such as the publication of bi-annual reports if one hundred or more complaints are received, and the issue of fining. Importantly, if the network does not designate a domestic agent or if the designated agent is not responsive, a fine of up to €5 million may be levied. For all other violations of the Act, including the negligent or intentional violation of its obligations, the social network may be fined up to €50million. This law is revolutionary, not because it seeks to regulate hate speech on social media and social media networks themselves, but because it adopts such a stringent approach to the regulation of hate speech on social media, as demonstrated through the obligation on networks to appoint agents solely for the purpose of this law, as well as the huge potential fines associated with the non-conformity to the legislation by the legislation.

Conclusion

The Code of Conduct between the European Commission and the IT Companies is an innovation in terms of regulating hate on social networks in the sphere of illegal hate speech. The results of the first monitoring period results were not positive in terms of how seriously YouTube, Twitter and Facebook took their commitments under this role. Although the second monitoring period and, to a greater extent, the third monitoring period did reflect an improvement in the commitment of social networks, there are still fundamental problems in relation to the role and enforcement of the Code of Conduct on Illegal Hate Speech. Firstly, although there is an improvement in the speed and removal of content by the third monitoring cycle, it must be underlined that the social networks were aware of the monitoring cycles and the organisations and persons involved in the process. We have no data in relation to how they respond to their duties under the Code when they are dealing with reports from users beyond a monitoring exercise. Secondly, as extrapolated on above, the enforcement of the Code of Conduct is absolutely reliant on the knowledge, will and intention of users. IT companies have to act upon hate speech only after it has been reported. Moreover, this Code of Conduct has been agreed

between the companies and the European Commission, with the Commission, of course, monitoring its implementation in relation to online hate in Member States but not beyond. So this is something for the EU. It is, nevertheless, a template that could be used by other countries and regions. Either way, regulation by social networks and the role of the Code of Conduct are definitely significant in the sphere of Internet hate regulation as social networks can remove material themselves and have a process by which this can be done. Moreover, through the amendment to the Audiovisual Directive to incorporate the issue of hate speech and through the German Act to tackle illegal content on social networks, which, although a national law, impacts other countries due to the borderless nature of the Internet, it is apparent that the legal regulation of online hate is going up on the European and some national agendas. A point to note here is that it was expected that Germany would be the first country to pass a regulatory law of this magnitude on illegal content online due to its traditionally restrictive position on hate speech. The above developments are definitely a light at the end of the Internet hate tunnel, where issues of multiple jurisdictions as well as technological realities, such as mirror sites and more, have resulted in the task of regulating hate online being more than a daunting one. Had it not been for the internal process of regulation on social networks which has been made semi-external given the role of the Commission, a neo-Nazi organisation's account which was blocked by Twitter would probably still be up and running and spreading hate as well as the account of a far-right member of the European Parliament who regularly tweeted homophobic statements.[72] For some, this could even be a positive thing according to the importance they attach to free speech and public discussion. If one endorses international human rights law and documents such as the International Convention on the Elimination of All Forms of Racial Discrimination, the Framework Decision on Racism and Xenophobia and the Additional Protocol to the Cybercrime Convention, such expression is unacceptable. However, hate speech is a contested topic with many commentators and even national legislators arguing in favour of free speech. The current regulatory process on social networks appears to be a middle ground for the different positions held in relation to what we should do with hate speech given that regulation on such networks, such as removing a post, is less severe than, for example, imprisonment. This position does not

[72] Anita Huslin, 'Twitter Blocks Offensive Accounts in Germany, U.K.; Deletes Tweets in France.': <http://www.npr.org/blogs/the two-way/2012/10/19/163243194/twitter-blocks-offensive/accounts-in-germany-u-k-deletes-tweets-in-france> [Accessed 1 May 2017]

mean that actions to implement the national criminal law cannot be taken. Either way, although the process of regulation by social networks is effective in tackling online hate, regulation in itself is not sufficient. Other measures need to be taken for purposes of tackling online hate in the long-term and for tackling online hate which does not meet criminal law thresholds. In relation to the last point, a central issue to hate found online, either on social networks or on other Internet platforms, is what happens to the majority of hate speech which is not deemed illegal by national law and/or by the Framework Decision on Racism and Xenophobia but still hurts people, groups and societies? As noted by the United Nations Special Rapporteur on Freedom of expression, there are three types of problematic expression, that which is a criminal offence under international law, that which is not a criminal offence but can result in restriction and civil suits and that which has no legal implications but still raises issues relating to respect and tolerance.[73] Assuming that type (a) and (b) can be dealt with by the law as well as social network regulation, type (c), albeit potentially being regulated by social networks, needs another solution. To this end, it is argued that social networks, as well as facilitating regulation of online hate through the procedures described above, also constitute positive platforms through which counter-narratives to hateful speech (illegal or not) can be developed. These can be done through commentary on material which is hateful but has not been removed, campaigns through groups and pages and more. As such, it is concluded that social networks provide a space in which haters can utter and disseminate their hate. At the same time, social networks also provide space for those who seek to respond to this hate and work on altering the rhetoric for purposes of establishing an equilibrium in terms of ideas and positions promoted online *vis-à-vis* minority groups (ethnic, sexual and more). Moreover, social networks, through their procedures of reviewing hateful material, as further enhanced by the Code of Conduct, are, in terms of infrastructure, ideal settings in which hate speech can be regulated.

[73] UNESCO 'Countering Online Hate Speech' (2015 UNESCO Publishing) 16: <http://unesdoc.unesco.org/images/0023/002332/233231e.pdf> [Accessed 1 May 2017]

BIBLIOGRAPHY

International Conventions/Treaties

Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems

Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law

European Convention on Human Rights

Explanatory Report to the Additional Protocol to the Cybercrime Convention

International Convention on the Elimination of All Forms of Racial Discrimination

Universal Declaration of Human Rights

Case Law

Delfi AS v Estonia, App. No 64569/09 (ECHR 16 June 2015)

Handyside v UK, Application no. 5493/72 (ECHR 1976)

Magyar Tartalomszol Galtatok Egyesulete and Index.hu ZRT v Hungary, App. No 22947/13 (2 May 2016)

Yahoo, Inc. v. La Ligue Contre Le Racisme et L'Antisemitisme, et al 145 F. Supp. 2d 1168, Case No. C-00-21275JF (N.D. Ca., September 24, 2001)

Books:

Fahy P. '*Achieving Quality with Online Teaching Technologies*' (eds. ERIC Clearinghouse 2000)

Tsesis A, '*Destructive Messages: How Hate Speech Paves the Way for Harmful Social Movements*' (eds. NUY Press 2002)

Book Chapters:

Vajić N & Voyatzis P, 'The Internet and Freedom of Expression: A Brave New World and The ECtHR's Evolving Case-Law' in Josep Casadevall, Egbert Myjer, Michael O'Boyle & Anna Austin (eds), *Freedom of Expression: Essays in Honour of Nicolas Bratza* (Wolf Legal Publishers 2012)

Journal Articles:

Alkiviadou N, 'Regulating Internet Hate: A Flying Pig?' (2016) 7 Journal of Intellectual Property, Information Technology and E-Commerce Law 3

Brennan F, 'Legislating against Internet Race Hate' (2009) 18 Information & Communications Technology Law 2, 123

Harris C, Rowbotham J & Stevenson K, 'Truth, Law and Hate in the Virtual Marketplace of Ideas: Perspectives on the Regulation of Internet Content' (2009) 18 Information & Communications Technology Law 2

Kiska R, 'Hate Speech: A Comparison Between The European Court of Human Rights and the United States Supreme Court Jurisprudence' (2012) 25 Regent University Law Review 107

Kubler F, 'How much freedom for racist Speech? Transnational Aspects of a Conflict of Human Rights' (1998) 27 Hofstra Law Review 2

Nemes I, 'Regulating Hate Speech in Cyberspace: Issues of Desirability and Efficacy' (2010) 11 Information and Communications Technology Law 3

Post R, 'Racist speech, Democracy and the First Amendment' (1991) 32 William and Mary Law Review 267

Shaw L, 'Hate Speech in Cyberspace: Bitterness without Boundaries' (2012) 25 Notre Dame Journal of Law, Ethics & Public Policy

Timofeeva Y, 'Hate Speech Online: Restricted or Protected? Comparison of Regulations in the United States and Germany' (2003) 12 Journal of Transnational Law and Policy 2

Wright G, 'Dignity and Conflicts of Constitutional Values: The Case of Free Speech and Equal Protection' (2006) 43 San Diego Law Review 527

Online Sources:

'A short history of the Council of Europe': <http://www.coe.int/T/E/Com/About_Coe/10_points_intro.asp> [Accessed 1 May 2017]

Facebook's Community Guidelines: <https://www.facebook.com/communitystandards#hate-speech>

Huslin A, 'Twitter Blocks Offensive Accounts in Germany, U.K.; Deletes Tweets in France.' Available at: <http://www.npr.org/blogs/the two-way/2012/10/19/163243194/twitter-blocks-offensive/accounts-in-germany-u-k-deletes-tweets-in-france>

Twitter statistics available at: <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>

Twitter's Terms of Service (Content): <https://twitter.com/tos?lang=en#usContent>

YouTube's Community Guidelines: https://www.youtube.com/yt/policyandsafety/communityguidelines.html

Other:

Code of Conduct on countering illegal hate speech online: <http://ec.europa.eu/justice/fundamental-rights/files/hate_speech_code_of_conduct_en.pdf>

Code of Conduct on countering illegal hate speech online: First results on implementation: <http://webcache.googleusercontent.com/search?q=cache:VckMt2f4jiEJ:ec.europa.eu/newsroom/document.cfm%3Fdoc_id%3D40573+&cd=1&hl=en&ct=clnk&gl=cy>

Council of Europe's Committee of Ministers Recommendation 97 (20)

Facebook statistics: <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>

General Recommendation No. 32 on the Meaning and Scope of Special Measures in the International Convention on the Elimination of Racial Discrimination (2009) CERD/C/GC/32

Report of the Special Rapporteur, Mr. Abid Hussain, submitted pursuant to Commission on Human Rights resolution 1997/26 (28 January 1998) E/CN.4/1998/40

Silva L, Mondal M, Correa D & Benevenuto F, 'Analyzing the Targets of Hate in Online Social Media' Proceedings of the Tenth International AAAI (Association for the Advancement of Artificial Intelligence) Conference on Web and Social Media (2016)
The Secretary-General, 'Preliminary Representation of the Secretary-General on Globalization and Its Impact on the Full Enjoyment of All Human Rights' paras 26-28, U.N. Doc A/55/342 (Aug 31 2000)

Twitter statistics: <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>

UNESCO 'Countering Online Hate Speech' (2015 UNESCO Publishing): <http://unesdoc.unesco.org/images/0023/002332/233231e.pdf>
Youtube statistics: <https://www.omnicoreagency.com/youtube-statistics/>